IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 4 (April. 2019), ||S (I) || PP 60-69

Low Level Feature Extraction Techniques in Content Based Image Retrieval: A Review

Manasha Saqib

Abstract: A database is a collection of information that is organized for easy storage, retrieval and update. This information can berepresented in many forms like text, image, chart, table, and graph etc. Here we focus on information in the form of images that is stored. Image Retrieval in the current scenario is the basic requirement of today's life. Due to the huge amount of different image types in the database from different sources for image recovery, different processing types are required to extract the relevant features from them. Techniques for Content Based Image Retrieval (CBIR) appeared in the 1990s. To describe image content, it uses low-level features such as color, texture and shape, and breaks through the limitation of traditional text query technique. Content Based Image Retrieval (CBIR) is a meaningful and increasingly popular approach that helps to retrieve image data from a large digital image database because it requires relatively less human interference.Content Based Image Retrieval (CBIR) is now a solution and source of accurate and rapid recovery. CBIR uses visual content to obtain relevant images from large databases in accordance with the interests of the user. The visual content (color, texture, shape, etc.) serves as the image characteristics. Features are ultimate interest measurements analyzed from a picture. It cannot be an amicable solution for accuracy and efficiency to use a single feature extraction for image retrieval. High-dimensional feature reduces the efficiency of the query, low-dimensional feature reduces the accuracy of the query, so that multiple features are better used for image recovery. The most important visual features are color, texture and shape.

Keywords: *Image Retrieval, Content Based Image Retrieval (CBIR), Color, Texture, Shape, Feature extraction.*

Date of Submission: 26-03-2019 Date of acceptance: 09-04-2019

I. INTRODUCTION

Due to the rapid development of internet technology, the image document has become an important source of information. Image recovery uses a computer system to recover images from a large database of digital images. It is difficult to recover certain images from all available images, so this problem image recovery system was developed to solve. As data and information are growing very rapidly, new techniques are needed and automated tool is generated that can intelligently help us transform the large amount of data into useful information and knowledge, so the new term data mining is introduced. Data mining refers to "mining" or extracting knowledge from large amounts of data. The other term that carries a similar or slightly different meaning to data mining is data mining knowledge, data / pattern analysis of knowledge extraction, etc. [1].

An image can be defined as a two-dimensional function, f (x, y), where x and y are spatial (plane) coordinates, and the amplitude of f is called the intensity or gray level of the image at that point at any pair of coordinates(x, y). If x, y and f's intensity values are all finite, the image is called a digital image [2]. Images have always been a part of human communication that is inevitable. Human beings always preferred concrete visual means (images, painting) to a greater extent to express ideas and convey information. The importance of Content-Based Image Retrieval (CBIR) is motivated by the growing hope of capturing images from the Internet from growing digital image databases. As the size of image databases increased exponentially, it became difficult to run large image databases, leading to research communities ' motivation to explore new algorithms for extraction of features.

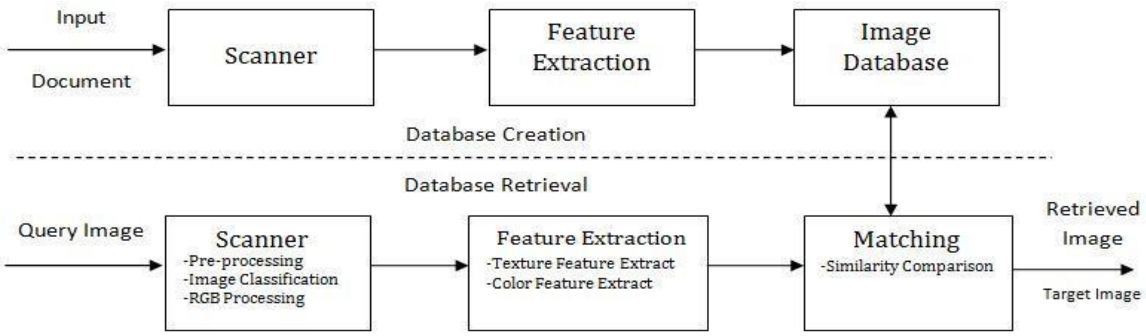

Fig.1 Database Creation And Retrieval [3]

Here fig.1 shows the phase of creation of the database and the phase of retrieval of the database. Using the query system, image is taken as input in the first phase, first scans the image through its content, features this extraction feature and stores it in the database of images. Another side is the retrieval phase of the database. In this phase query image is taken as input, scan the image using different techniques and methods to improve its quality such as pre-processing, classification of images and processing of RGB. Next stage is the extraction of features, for example. Extraction of color, extraction of texture and extraction of shape. The last stage is to match the existing feature, store it in the database if it matches, the image being retrieved is the target image of the call.

II. OBJECTIVE

In many fields, CBIR was a very important and effective research area. Increased bandwidth availability will help increase user usage of the internet in the future by searching images from large database and giving appropriate results. Hence, a significant challenge that needs to be addressed is to quickly retrieve images from large databases. Image retrieval system searches for large database images and attempts to find accurate or almost identical images. CBIR can significantly improve the accuracy of the returned data and is effective for conventional text-based image search. To distinguish an image from other images, color, texture, histogram and shape features are used. The objective of the paper is to propose a new CBIR system; an important system task is 1) to reduce the "semantic gap" between low-level image characteristics and the richness of human semantics and 2) to reduce the overall time of retrieval. CBIR research and development issues cover a range of topics, many of which are shared with mainstream image processing and retrieval of information. Some of the goal may be to extract color, texture, shape and histogram characteristics from images, provide compact storage for large image databases, match query and stored images in a manner that reflects judgments of human similarity.

III. LITERATURE REVIEW

[4] In 1992, the United States National Science Foundation organized a workshop on the Visual Information Management System to identify a new direction in the image database management system, introducing the term Content Based Image Retrieval, highlighting the use of color and shape as the most important feature extraction criteria for the image recovery system. Since then, the CBIR has been adopted to describe a process of image retrieval.

S. Mori Tamura and T. Yamawaki [5] in 1973, proposed texture representations based on human perception psychological studies and consisting of six statistical characteristics, including coarseness, contrast, directionality, regularity, line-like characteristics, and roughness to describe different texture properties. Extraction tamura properties are very meaningful in the context of texture features, and these benefits make tamura features in Texture Based Image Retrieval.

H. Zhao, Z. Xu and P. Hong[6] proposed textural extraction based on coarseness in 2009.To improve the performance they used textural coarseness and compared the result with the Gray Level Co-occurrence Matrix textural coarseness, Fractual dimension textural coarseness and tamura texture model. And among the three, they proposed the performance of the textured tamura model describing coarseness is best followed by the other two methods.

H. Yao and B. Li [7] in 2003, proposed a retrieval system using Gray Level Co-occurrence Matrix and sobel detector applied, edge detection by texture segmentation is one of the methods since, by considering only texture properties such as coarseness energy, and some of the information is loss.They therefore proposed the combination of both texture segmentation method edge detection and texture properties, obtained the high precision value recovery.

P. Gangadhara Reddy [8] in 2010, used feature extraction and proposed Color Co - occurrence (CCM) matrix based on Gray Level Co - occurrence Matrix extract features from any color plane for each plane and proposed image retrieval based on GLCM and color multi - fusion, and retrieved texture features based on color space HSV based image retrieval using similarity measures such as Euclidean distance. Proposed features of CCM textures and CCM color composition enhance the image retrieval performance, which is the major research value.

N. Chaturvedi, S. Agrawal and P. Kumar Johari [9]proposed CBIR based on contrasting texture characteristics, coarseness, and statistical characteristics of directionality. They first suggested extracting the texture-based feature vectors from the query image, then applying the similarity measurement algorithm to the extracted feature vector from which relevant images from the database are extracted.

Chih-Chin Lai and Ying-Chuan Chen [10] proposed an interactive genetic algorithm (IGA) to reduce the gap between the results of the recovery and the expectations of the users, known as the semantic gap. They used HSV color space that matches human perception of colors and separates the luminance component from chrominance components. They also used texture characteristics such as entropy based on the co-occurrence gray level matrix and the edge histogram.They compared this method with other approaches and better results were obtained.

RishavaChakravarti, Xiannong Meng [11] in paper The Color Histogram Based Image Retrieval used the technique of color histogram to retrieve the images. This method allows the retrieval of images transformed in terms of their size and translated by rotations and flips.

A.Ramesh Kumar, D.Saravanan [12] in 2013, proposed a content-based image recovery (CBIR) using the technique of color histogram. And a color histogram is a representation of a picture's color distribution. A color histogram represents the number of pixels with colors in each of a fixed list of color ranges that span the color space of the image, the set of all possible colors, for digital images.

Rajshree S. Dubey, Niket Bhargava and Rajnish Choubey [13] illustratedthe image mining methods depending on the color Histogram, the image's texture. The query image will be considered, then the COLOR Histogram and Texture will be created and the resulting image will be found accordingly. The calculation time for RGB color space is not considered in this approach.

M.Babu Rao, Dr. B.Prabhakara Rao and Dr. A.Govardhan [14] proposed an efficient image recovery technique using a picture's dominant color and texture features. The method proposed yielded higher average accuracy and average recall with reduced vector dimension feature.

IV. CONTENT-BASED IMAGE RETRIEVAL

Content Based Image Retrieval System (CBIR) was introduced in early 1990.Content Base Image Retrieval System is image retrieval techniques based on different features such as color, texture and shapes. Also known as content-based image recovery system (CBIR) means the search will analyze the actual image content. In CBIR, an image's visual content is automatically extracted. There are many features that make an image; but four of them are regarded as the main features, i.e. color, texture, shape and spatial properties. The retrieval of images depends entirely on these characteristics.However, implicit consideration is given to spatial properties. Color, texture and shape are the main features to consider. CBIR system has the advantage that it is a fast method and can automatically extract the low-level feature. This low-level feature such as color, texture, image-extracted shape to measure the similarity between different images and to retrieve similar images.

The features of each image are efficiently extracted and stored in the database in a typical CBIR process. Extract the corresponding features from the query image to retrieve the images and search the image database to identify the similar images and return the results [15][16] Thus, a typical CBIR system (Fig.2) consists of three major components and their variations depend on the features used.

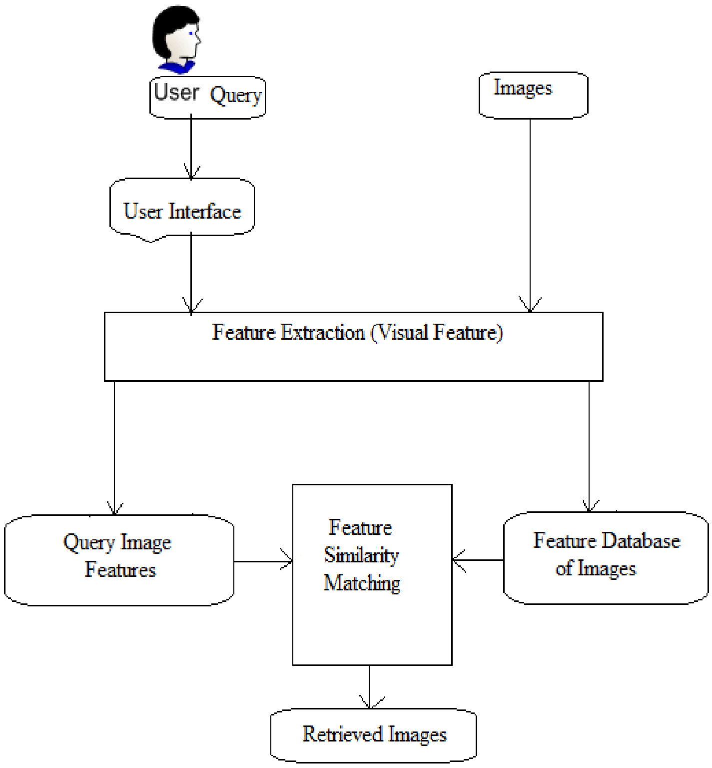

Fig.3: Content Based Image Retrieval [17]

1. **Feature extraction** – Analyzing raw image data to extract feature-specific information plays an important role in supporting efficient and rapid recovery from image databases of similar images.
2. **Feature storage** – Provide efficient storage of extracted information, as well as help improve search speed.
3. **Similarity measure** –Measure the difference between images to determine the relevance between images, resulting in a visually similar result.

Extraction of features is crucial throughout the CBIR process. Extraction of characteristics and measure of similarity are very dependent on the characteristics used.

CBIR considers two main factors as-

In the query, it uses computable image and video properties such as color, shape, texture, object movement and other graphical information.

Graphical Query Languages is a query that involves drawing, selecting and retrieving digital image graphical features.

V. FEATURES EXTRACTION

A feature is defined as an interesting part of an image, and many computer vision algorithms use features as a starting point. Because features are used for subsequent algorithms as the starting point and main primitives, the overall algorithm will often be just as good as its feature detector. Extraction of features is the core of image retrieval based on content. As we know, in most computer vision tasks, raw image data cannot be used directly. First of all, two reasons behind this, the image's high dimensionality makes it difficult to use the entire image. A lot of the information embedded in the picture is redundant as well.

Therefore only an expressive representation of the most important information should be extracted instead of using the whole image. The process of finding the expressive representation is called the extraction of the feature and the resulting representation is called the vector of the feature. Feature extraction can be defined as mapping the image from the space of the image to the space of the feature. Now days, it is still a difficult task to find good features that well represent an image. The content of images can distinguish between the content of visual and semantic content. Usually features represent the visual content. Furthermore, visual content can be divided into general or specific domains. For instance, the features that can be used to search would represent general visual content such as color, texture, and shape. On the other hand, the features used to search human faces are domain-specific and may include knowledge of the domain. In pattern recognition literature, the domain-specific features are better covered. There is no single best presentation for a given feature due to subjectivity of perception. There are multiple representations for any given feature that characterize the feature from different perspectives.

When we talk about an image's semantic content, it's not easy to extract. Annotation and/or specialized visual content-based inference procedures also help to some extent to obtain semantic content. The main point for selecting the features to be extracted should be guided by the following concerns:

The features should contain sufficient information about the image and should not require any domain-specific knowledge, and it should be easy to calculate so that the approach is feasible for large image collection and rapid recovery. Another thing is that it should be related well to the human perceptual characteristics as users finally determine the suitability of the image retrieved.

Extraction of features plays an important role in the Image Retrieval system and improved feature selection gives greater precision. Feature extraction essentially separates the visual information from the image and stores it in a feature database in the form of feature vectors. This feature value (or a set of values) called image feature vectors finds the image information from the extraction of the feature. To compare the query image with the images stored in the database, these feature vectors are used. Features can be characterized as a way of distinguishing one class of object from another in the field of pattern recognition. When extracting features in CBIR, the most important problem is obtaining the most relevant images in the selection of features. When designing an image recovery system, each feature can have several representations and different representations aspects of the feature.

VI. Color Feature Extraction Techniques

The color feature is a very basic visual feature that is used in image recovery based on image color similarity. Background issues are relatively robust and are not dependent on changes in image size, orientation, and scale. Most of the images are distinguished based on human color characteristics. The different techniques of extraction of color features are given as follows:

1. Color Histogram

Because of the stability and robustness, color descriptor has been widely used for image recovery in CBIR. It is not changed due tothe translation, rotation and scale changes of image. Most commonly used method is the color histogram and simple to implement. The color histogram defines the three color channel intensity probability [18]. The image histogram is taken by counting the number of each color pixel in the image and containing each color pixel in separate bins. The image histogram is not changed to the image plane's rotation and translation, and is slowly changed when the angle view is changed [19]. The Color Histogram can be represented as $h_{A, B, C}$ (a, b, c) = N.Prob (A=a, B=b, C=c), where N is the number of pixels in the image and A, B and C are the three color channels.

2. Color Correlogram

A colorCorrelogram provides information on how distance changes the pairs of colors[20]. The color Correlogram is not a method of refining the histogram or a method of partitioning the image. The color Correlogram features are highlighted as:

- It describes the color correlation in the spatial plane,
- It is used by local spatial color correlation to describe the global distribution.
- It is relatively simple in computing, and

The size of the feature is fairly small [21]. A color Correlogram of an image is a table obtained by calculating the number of color pixels of j at a distance k from the color ithe kth entry at location (i, j), divided by total number of pixels in the images.

3. Dominant Color Descriptor

The Dominant Color Descriptor (DCD) describes the typical colors in an image or image region [22]. It is used to obtain similar images from the database and to browse the database of images based on single or different color values. Compared to conventional histogram-based descriptors, the DCD can provide the powerful and compact salient color representation. The DCD is defined as F= $\{\{c_i, p_i, v_i\}, s\}$, (i=1, 2 ... N), where N is the dominant color number [23]. Color value c_i is a vector of the corresponding component values for the color space. The percentage pi is the pixel fraction corresponding to color ci in the image or image region, and variance describes the color values variance.

4. Color Co-occurrence Matrix

The color co-occurrence matrix (CCM) is a common method used to capture image color variations that gives the color characteristic. It is used to calculate the likelihood of occurrence between each pixel and its adjacent pixel of the same pixel color [24]. Each pixel in the image corresponds to the four adjacent pixels so that each image can be represented by four scan pattern image motifs, which can be further built into four two-dimensional matrices. The scan pattern motifs are generated from these four matrices to capture the image color variation.

VII. TEXTURE FEATURE EXTRACTION

Texture is a very important feature in analyzing many types of images that appear in nature everywhere, such as natural images, remote sensing images and medical images [25]. Texture can be defined as the superficial phenomenon of natural objects ' human visual systems. Texture can be attributed to nearly everything in nature and also incorporates its texture structure of any image.Texture can be attributed to nearly everything in nature, and the texture structure of any image also incorporates repeated patterns of most parts. Texture is generally referred to as' texels.' All can recognize texture, but it's not easy to define. Texture does not take place over a point, but rather over a region. By quantitative and qualitative analysis, texture can be analyzed.

However, texture can be considered as repeated patterns of pixels over a spatial domain, resulting in textures that may appear random and unstructured by adding noise to patterns and their repetition frequencies. Texture properties are the visual patterns in an image that have homogeneity properties that are not the result of a single color or intensity being present. For example, the different texture properties perceived by the human eye are regularity, directionality, smoothness and coarseness.

VIII. TAMURA TEXTURE FEATURE

One of the first descriptions given by the Tamura [26], according to quantitative analysis, proposed six textural properties and gave common descriptions of all texture patterns. Tamura gives six different texture characteristics, i.e. Coarseness, contrast, directionality, line similarity, regularity and rigidity.

- **Coarseness**

Coarseness essentially refers to the distance of spatial variations in gray levels, which is implicitly related to the size of the texture-forming primitive elements. It has the direct relationship to the rates of scale and repetition and the most basic feature of the texture. An image will contain repeated patterns of textures at different scales, coarseness aims to identify the largest size where there is a texture, even where there is a smaller micro texture.

- **Contrast**

Contrast measures the distribution of gray levels in an image that varies and to what extent is biased to black or white in its distribution. The contrast is defined by the second order and standardized fourth - order central moments of the gray levels.

$$Contrast=\sigma / (\alpha 4)$$

$$\alpha 4=\mu 4/\sigma 4 \quad (3)$$

Where, $\mu 4$ is the fourth moment about the mean and 2 is the variance. n=1/4 to give the closest value.

- **Directionality**

Directionality of an image is measured against its directional angles by the frequency distribution of oriented local edges. Over a region, it is a global property. This texture feature given by tamura does not distinguish between orientations or patterns, Directionality measures the total degree of directionality in an image. It is the most important matrix feature given by tamura to distinguish how uniform the region is from another image.

- **Line-Likeness**

Line-likeness in an image is an average coincidence of edge direction that co-occurred in pixel pairs separated by a distance in each pixel along the edge direction.

- **Regularity**

Regularity measures the pattern that occurred regularly or similarly in the image.

$$F\ (regularity) = 1 - r(S_{(crs)} + S_{(con)} + S_{(dir)} + S_{(line)})$$

- **Roughness**

Roughness is the summation of measurements of contrast and coarseness

Roughness= Contrast + Coarseness

In most cases, only the first three features are used for CBIR system because these features capture a texture's high-level perceptual attributes and are also useful for image browsing. Textures of images have useful image processing and computer vision applications. These include: recognition of image regions using texture properties, known as texture classification, recognition of texture boundaries using texture properties, known as texture segmentation, texture synthesis, and generation of texture images from known texture models. Since there is no accepted texture mathematical definition, many different methods have been proposed over the years for computing texture features. Unfortunately, with all types of textures there is still no single method that works best.

Haralick Texture Feature

Gray Level Co-occurrence Matrix (GLCM) is a statistical method for examining texture features that takes into consideration the pixel spatial relationship, also known as Gray Level Spatial Dependence. In this a

GLCM matrix is created by calculating how often in a specific spatial relationship to a pixel with the value j a pixel with the intensity value i occurs. GLCM consists of frequencies at which a certain vector occurs in the image separating two pixels. GLCM properties depending on the distance and angular or directions such as horizontal, vertical, diagonal, anti-diagonal relationship between the pixels. Many statistical characteristics of texture in an image are based on the matrix of co-occurrence representing the relationship of the second order of gray level pixels in an image. Different statistical and information theoretical properties of co-occurrence matrices can be used as textural features and the limitation with these features is expensive to calculate, and they were not very efficient for classification and retrieval of images.

Haralick [27] proposed 28 types of textural features, each extracted from the Co-occurrence Matrix of Gray Level. Assume that an image input has M total number of pixels in horizontal direction and M total number of pixels in vertical direction. Assume that the gray level appearing at each pixel is quantified to z number of levels, assume that Nx= 1, 2, 3 ... M consists of horizontal space and Ny=1, 2, 3....... N consists of vertical space and G= 0,1,2,3... Z consists of a set of Z quantified gray levels. The Gray Level Co-occurrence matrix is calculated in a given distance d and direction by using gray scale pixels i and j, expressed in different directions as the number of co-occurrence matrix.

$P(i, j|d, \theta) = p(i, j|d, \theta) / \sum_i \sum_j p(i, j|d, \theta)\ ij$ (6)

Among these, five features are Contrast, Correlation, Entropy, Energy and Homogeneity

Contrast

Contrast measures intensity over the entire image between a pixel and its neighbor and is considered to be zero for constant image and is also known as variance and moment of inertia.

Contrast = $\sum_{i,j} (i - j)^2 p(i, j)$ (7)

Correlation

Correlation measures how pixels over the entire image are correlated to their neighbors.

$Correlation = \sum_{i,j} (i-\mu_i)(j-\mu_j)\ p(i, j)/\delta_i \delta_j$

Entropy

Entropy provides image complexity measurements and this complex texture tends to be more entropic.

$Entropy = \sum i \sum j P(i,j)$

Energy

Energy is the sum of square elements in the GLCM and, by default, is one for constant image.

$Energy = \sum_{i,j} (i, j)^2$

IX. SHAPE FEATURE EXTRACTION

The shape is one of the common features used in CBIR systems. The image retrieval based on shape is the measurement of similarity between shapes represented by their characteristics. For image content description, shape is an important visual feature and one of the primitive features. An object's shape is the characteristic surface configuration that the outline or contour represents. Shape recognition is one of the ways of executing human perception of the environment. It is important in CBIR as it corresponds to image region of interest.Representations of the shape feature are categorized by the techniques used. They are based on boundaries and region [28]. All the pixels within a shape are taken into account in region-based techniques in order to obtain the representation of the shape. Common region-based methods to describe shape using moment descriptors [29]. Region moment representations interpret a standardized gray level image function as a 2-D random variable probability density. Hu [30] gives the first seven invariant moments, derived from the normalized central moments of the second and third order. Since moments combine information across an entire object rather than just providing information at a single boundary point, they capture some of the global properties that are missing from many pure representations based on contours.

X. HU-MOMENT SHAPE FEATURES

In 1962, Hu-Moment [31] proposed seven associated region properties invariant to rotation, scaling and translation (RTS) and also known as Algebraic Moment Invariants. Invariants of the moment calculated from each window are used to form vectors of features. They define simple, set region properties that can be used for class identification and also shape identification, and Hu originally proposed this classic technique for algebraic invariant generation.

Suppose R is an image, p+q, central moments or R forms as

$\mu_{p,q}, =\Sigma_{x,y,y} (x-x_c)^p (y-y_c)^q$

(x_c, y_c) is the center of object .For scale –independent nature, central moments can be standardized as

$\eta_{p,q}=\mu_{p,q}/\mu 0^2, 0$

$\gamma=p+q+2/2$

Based on these moments, Hu bring forward seven moments independence of translation, rotation and scaling.

$\emptyset 1= \mu_{2,0}+\mu_{0,2}$

$\emptyset 2= (\mu_{2,0}-\mu_{0,2})^2+4\mu^1_{1,1}$

$\emptyset 3= (\mu_{3,0}-3\mu_{1,2})^2+ (\mu_{3,0}-3\mu_{2,1})^2$

$\emptyset 4= (\mu_{3,0}+\mu_{1,2})^2+ (\mu_{0,3}+\mu_{2,1})^2$

$\emptyset 5=(\mu_{3,0}-\mu_{1,2})(\mu_{3,0}+\mu_{1,2})[(\mu_{3,0}+\mu_{1,2})^2-3(\mu\_{2,1}+\mu\_{0,3})^2]+(3\mu_{2,1}-\mu_{0,3})(\mu_{2,1}+\mu_{0,3}).[3(\mu_{3,0}+\mu_{1,2})2-(\mu_{2,1}+\mu_{0,3})^2$

$\emptyset 6=(\mu_{2,0}-\mu_{0,2})[(\mu_{3,0}+\mu_{1,2})2-(\mu_{2,1}+\mu_{0,3})^2]+4\mu_{1,1}(\mu)(\mu_{2,1}+\mu_{0,3})$

$\emptyset 7=(3\mu 2{,}1-\mu 0{,}3)(\mu 3{,}0+\mu 1{,}2)[(\mu 3{,}0+\mu 1{,}2)2-3(\mu 2{,}1+\mu 0{,}3)2]-(\mu 3{,}0-\mu 1{,}2)(\mu 2{,}1+\mu 0{,}3)[3(\mu 3{,}0+\mu 1{,}2)2-(\mu 2{,}1+\mu 0{,}3)2]$

Ø7, is the skew moment, and this skew invariant is useful in distinguishing mirror images. This moment used in extractions of features can be generalized in order to achieve pattern identification not only independently of position, size and orientation but also independently of parallel projection. Contour-based shape representation is more popular when compared with region-based shape representation. Representation of shape based on contour only takes advantage of shape boundary information. Simple shape descriptors based on contour include area, perimeter, compactness, eccentricity, elongation, and orientation. Fourier descriptors, grid descriptors, and chain codes [32] are complex boundary-based descriptors.

One of the most useful methods used in image processing is contour detection. Generally, the contour can be connected to a rapid gray level change in the observed image in digital images. The basic idea behind this work is to use the model of centroid-radii to represent forms. In this method, radii lengths of the shape are used to represent the shape from boundary to centroid. If the shape is used as a feature, the first step to extract that feature could be edge detection. The canny edge detector is used in this work to determine the edge of the object in the image [33].Edge extracted from an image tells about the entire content of an image. There are different techniques of edge detection i.e. Prewitt method, Sobel method, canny method etc.

On the bases of region-based shape extraction, boundary-based shape extraction, and contour-based shape extraction, there are different methods for extracting image shape. Compared to other extraction features, shape extraction is quite complicated as it required different transformations such as scaling, shifting, rotating, etc. to be applied to the image to capture the exact shape of the image. By identifying the corners in the image, the chain code method is used to determine the boundary of an image. Shape is a key attribute of segmented image regions, and its effective and robust representation plays an important role in the retrievalprocess. The way in which such representations are matched is synonymous with shape representation. Horizontal distance vector describes the object's shape variance from top to bottom, where the vertical distance vector describes the object's shape variation from left to right.

In a horizontal segmentation image is divided into horizontal segments and traces the coordinate points and determines the shape of the object because the sitting location problem of the object cannot be the same as in a vertical segmentation image divided into vertical segments. The chain code method calculates the binary image boundary, the object area represents the number of pixels within the closed boundary of the binary image, and the horizontal and vertical distances are represented by calculating the distance between boundary lines.

XI. SIMILARITY MEASUREMENT

An image may contain visual or semantic information. Visual information can be represented in shape, color, texture, and spatial relationships. The visual characteristics extracted are considered as a vector of features that are kept in the database of features. Different similarity measurement techniques that match the similarity between the query image and the stored image in the database perform the retrieval process. The query result may not be a single picture. It is a series of images categorized with the query image by the similarity of the image retrieved. Similarity measures affect the retrieval performance of CBIR systems. Measuring how close a vector to another vector is referred to as measuring similarity [34].In order to find the correspondence of the query image with the image stored in the database, similarity measures are used. For image retrieval, many similarity measures were developed based on empirical estimates of feature distribution. Different similarity or distance measurements will have a significant impact on the retrieval performance of an image retrieval system.We denote I=Query image J=Image in database fi (I) = represent the number of pixels in, ith ^{bin} of query image.

XII. EUCLIDIAN DISTANCE

Due to its higher accuracy and effectiveness, the Euclidian Distance [35] matrix is mostly used for similarity measurement in contextual image retrieval from database. By calculating the square root of the sum of the square absolute differences, it measures the distance between the two feature vectors of images and is calculated and denoted by ED.

$$Euclidean\ Distance = d = \sqrt{\sum_{i=1}^{N} \left(Xi - Yi \right)^2}$$

XIII. PERFORMANCE MEASURES

A key issue in Content-Based Image Retrieval (CBIR) is the evaluation of retrieval performance. The performance measurement used in image recovery borrows from the field of information retrieval and is based on two primary metric figures that are precision and recall. Precision is the number of documents that are retrieved. Recall is the number of relevant documents to be retrieved in the database [36].

Precision = No. of relevant images in the retrieved images/ No. of the retrieved image

Recall = No. of relevant images in the retrieved images/ No. of relevant images in database

Precision can be interpreted as a measure of exactness, whereas recall provides a measure of completeness. A perfect 1.0 accuracy score means that every image retrieved is relevant, but it does not give any insight as to whether all relevant documents are being retrieved. A perfect 1.0 recall score means all relevant images are retrieved but it does not say anything about how many irrelevant images might have been retrieved as well.

XIV. CONCLUSION

Image retrieval based on content is growing technique in the processing of images from different decades. CBIR has different techniques to retrieve suitable images from a large database. The best results are achieved by the CBIR techniques that use shape and layout together with color and texture. Selecting features is an important aspect of image retrieval in order to capture the image that is most relevant to user interest. High recall values obtained are required to compare the retrieved instances with the actual (relevant) instances. The basic concept is that without the laborious task of typing keywords, we can use input as an image and retrieve required images based on shape, color and texture features. Variations in the extraction methodologies of features have been found to ensure better and more accurate retrieval of relevant images from the large database. There is currently a lot of research being done to improve the methods of extraction of features. The ultimate goal is to achieve higher retrieval efficiency from large database of images by improving the speed, efficiency and accuracy. In the field of content-based image retrieval, there is still much research going on for its faster and more accurate behaviour.

REEFERENCES

[1]. J. Han and M. Kamber, Data mining: concepts and techniques.

[2]. R. C. Gonzalez and R. E. Woods, *Digital image processing*.

[3]. M. Sahu, M. Shrivastava and M.A.Rizvi, “Image mining: A new approach for data mining based on texture,”*ICCCT '12 Proceedings of the 2012 Third International Conference on Computer and Communication, IEEE* pp. 7-9, 2012.

[4]. T. Kato, “Database architecture for content- based image retrieval”, Image Storage and Retrieval Systems, Proceedings of SPIE- The International Society for Optical Engineering1662, pp. 112-123, April 1992.

[5]. S.M.Tamura and T. Yamawaki, “Textural Features Corresponding to Visual Perception”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. smc-8, no. 6, pp. 460-473, June1978.

[6]. H. Zhao, Z. Xu and P. Hong, “Performance evaluation of three classes of textural coarseness”, IEEE, 2009.

[7]. H. Yao and B. Li, “An efficient approach for texture-based image retrieval”, *Neural Networks and Signal Processing*, vol. 2, pp. 1039-1043, 2003.

[8]. P. G. Reddy, “Extraction of Image features for an Effective CBIR System”, IEEE, pp. 138-142, 2010.

[9]. N. Chaturvedi, S. Agrawal and P. K. Johari, “A novel Approach of image retrieval based on texture”, *International Journal of Enhanced Research in Management & Computer Applications*, vol. 3, no. 1, pp. 42-48, January 2014.

[10]. C. Lai and Y.C. Chen,“A User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm*”, IEEE Transactions on Instrumentation and Measurement,* vol. 60, issue no. 10, October, 2011.

[11]. R. Chakravarti and X, Meng "A Study of Color Histogram Based Image Retrieval", *International Conference on Informational Technology*,2009.

[12]. A.R.Kumar, D.Saravanan "Content Based Image Retrieval Using Color Histogram", *International Journal of Computer Science and Engineering (IJCSIT*), vol.4, Issue-2, pp. 242-245, 2013.

[13]. R.S. Dubey, N.Bhargava and R. Choubey,"Image Mining using Content Based Image Retrieval System", *International Journal On Computer Science and Engineering*, vol. 02, no. 07, pp. 2353-2356, October 2010.

[14]. M.B.Rao, Dr. B.P.Rao, Dr. A.Govardhan, "Content based image retrieval using Dominant color and Texture features", *International Journal of Computer science and information security*, Vol.9 issue No. 2, pp.41 – 46, February 2011.

[15]. R.Juan, "Intelligent image retrieval technique based on content,"*Lanzhou University of Technology*, pp 32-64, 2006.

[16]. A.K. Jain, J.E. Lee, R. Jin and N. Gregg, "Content based image retrieval: an Application to tattoo images*", IEEE International Conference on Image Processing*, pp. 2745-2748, November 2009.

[17]. S.Lata and P.P.Singh, "A Review on Content Based Image Retrieval System,"*International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, May 2014.

[18]. S. Murala, A. B. Gonde, and R. P. Maheshwari, "Color and Texture Features for Image Indexing and Retrieval," *2009 IEEE International Advance Computing Conference*, 2009.

[19]. R. Ferrari, R. Rangayyan, J. Desautels, and A. Frere, "Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets," *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 953–964, 2001.

[20]. N. Cressie, G. Upton, and B. Fingleton, "Spatial Data Analysis by Example: Volume 1: Point Pattern and Quantitative Data," *Technometrics*, vol. 29, no. 1, p. 114, 1987.

[21]. J. Huang, S. Ravi Kumar, M. Mitra, W. Zhu and R. Zabih, "Image indexing using color Correlogram,"*1997 IEEE, Computer Society Conference*, pp. 762-768, 1997.

[22]. B. Rao and A. Govardhan, "CTDCIRS: Content based image retrieval system based on dominant color and texture features," *International Journal of Computer Applications 18*, vol. 6, pp. 40-46, 2011.

[23]. J. Ohm, L. Cieplinski, H. Kim, S. Krishnamachari, B. S. Manjunath, D. S. Messing and A. Yamada, "The MPEG-7 color descriptors," *IEEE Transactions on Circuits and Systems for Video Technology,* 2001.

[24]. C. Lin, R. Chen and Y. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image and Vision Computing 27*, vol. 6, pp. 658-665.

[25]. Sukhwinder Singh, "A Closure Looks to Iris Recognition System", *International Organization of Scientific Research Journal of Engineering*, vol. 3, issue no.3, March, 2013.

[26]. S. Mori Tamura and T. Yamawaki, "Textural Features Corresponding to Visual Perception", *IEEE Transactions on Systems, Man and Cybernetics*, vol. smc-8, no. 6, pp. 460-473, June 1978.

[27]. K. R. M. Haralick, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. smc-3, no. 6, (1973) November, pp. 610-621.

[28]. R. Datta, J. Li, and J. Wang, "Content-based image retrieval - approaches and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, Article 5, pp. 1-60.

[29]. C. Teh and T. Roland, "On image analysis by the methods of moments," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496-513, 1988.

[30]. M. Hu. "Visual pattern Recognition by Moment Invariants," *IEEE Transactions on Information Theory, IT*, vol. 8, pp. 179-187, Feb. 1962.

[31]. Hu, "Visual Pattern Recognition by Moment Invariants", *IRE Transactions on Information Theory*, pp. 179-187, 1962.

[32]. M. Lew, N. Sebe, C. Djeraba and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia*.

[33]. Computing, Communications and Applications, vol. 2, no. 1, pp. 1-19, February 2006.

[34]. R. Chaudhari, A. M. Patil, "Content Based Image Retrieval Using Color and Shape Features", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol.1, 2012.

[35]. C. Chin Chen and H. Ting Chu, "Similarity Measurement between Images", 2005.

[36]. N. Puviarasan, Dr. R. Bhavani, and A. Vasnthi, "Image Retrieval Using Combination of Texture and Shape Features", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, pp. 5873-5877, 2014

[37]. M. R. Mustaffa, F. Ahmad, R. W. O. Rahmat, and R. Mahmod, "Content-Based Image Retrieval Based On Color-Spatial Features," *Malaysian Journal of Computer Science*, vol. 21, no. 1, pp. 1–12, 2008.